Jessica Lim 31954081

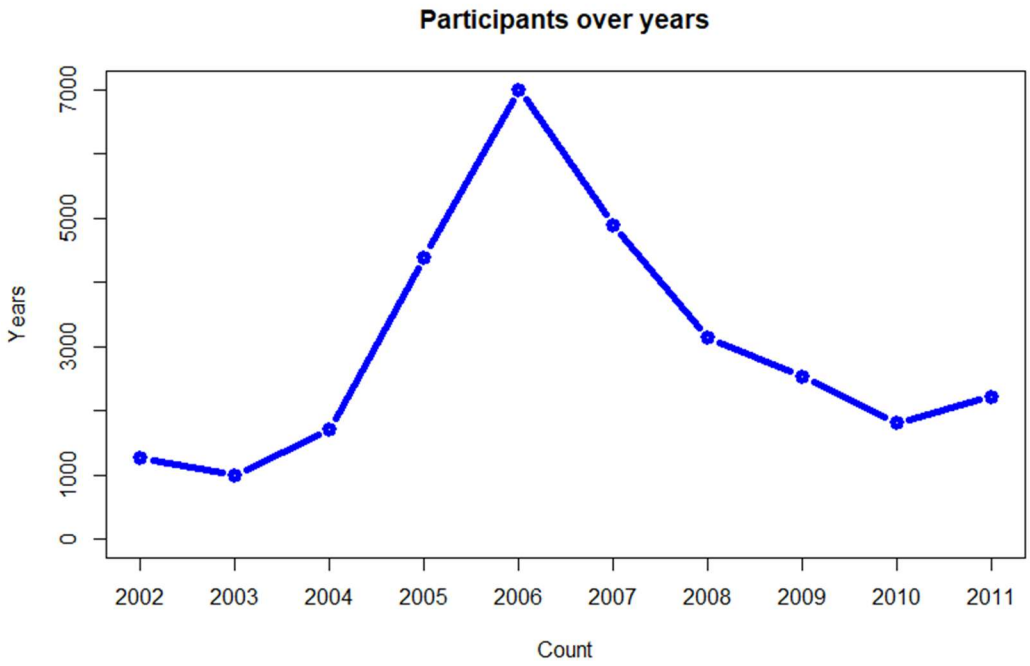
FIT3152:
Assignment 1

Analyze the activity, language use and social interactions of an on-line community using metadata and linguistic summary from a real on-line forum and submit a report of your findings

By Jessica Lim, Monash University Malaysia, 1st May 2022

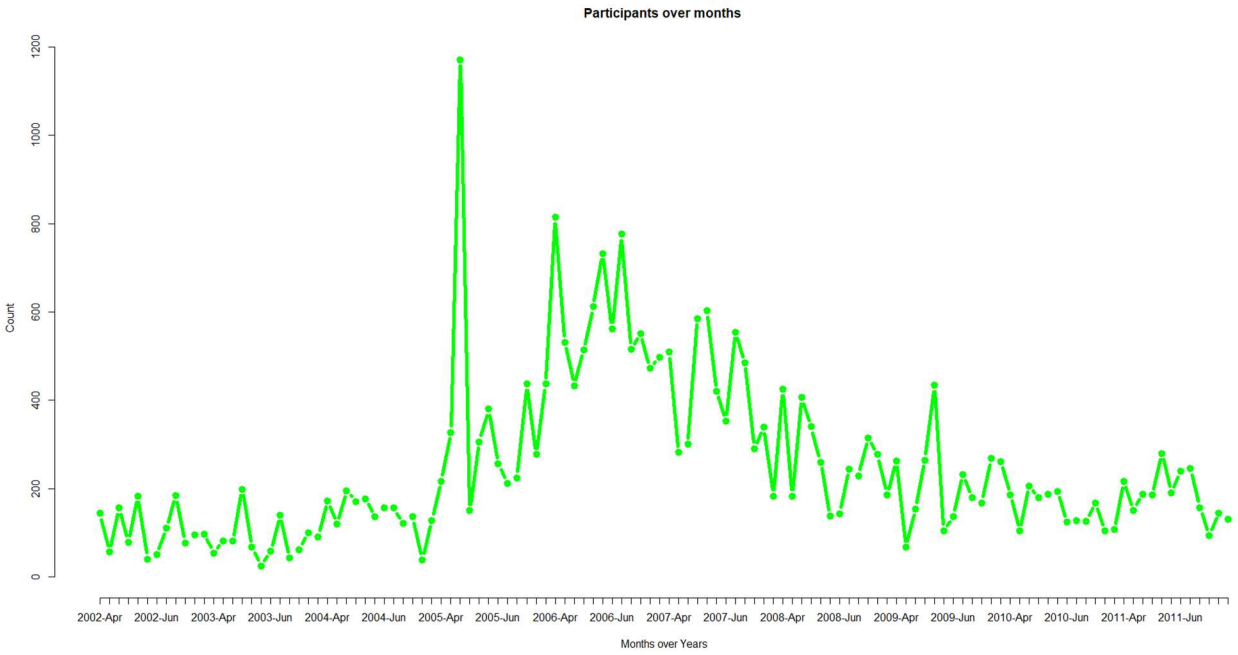
A.1. How active are participants over the longer term (that is, over months and/or years)? Are there periods where activity increases or decreases? Is there a trend over time? (3 Marks)

For this question, my definition of how active the participants are seen by how much are the posts. And so i achieved this graph by filtering all of the data by years and months.



From the data achieved by the visualization of the number of post over the years. We can see that there is a significantly rise from 2004 (after reaching the lowest count in 2003) to 2006 being the highest peak of most active. After the highest peak from 2006, there is a significant drop until 2010 in which rise for a bit again on 2011.

But of course to see the data more clearly i tried to visualize the data by months as we can see the graph below



As we can see in this graph same like the yearly graph. We can see that in total 2006 acumulate for the most activity however when we see in this graph the highest peak instead of in 2006 it shows that the peak of each month is in 2005 in April or May approximately. And from this graph we can also see that the graph line going up and down from April 2002 to the end of 2011.

So, in conclusion, there is a constant trend of up steadily going up then going down. The possibility of this happen is that because in 2003 this forum might be still unknown and in 2005/2006 this forum may be a hit but probably after 2006 there might be a new more interesting web forum or people might start getting bored with the forum.

A.2. Looking at the linguistic variables, do the levels of these change over the duration of the forum? Is there a relationship between linguistic variables over the longer term? (3 Marks)

Below are the correlations over all variables except for time and date filtered for every year. As we can see it shows a really varies color and plot this shows that there are some changes of the strongest correlation in over the years. The interpretation of correlation is:

-1: Perfect negative correlation. The variables tend to move in opposite directions (i.e., when one variable increases, the other variable decreases).

0: No correlation. The variables do not have a relationship with each other.

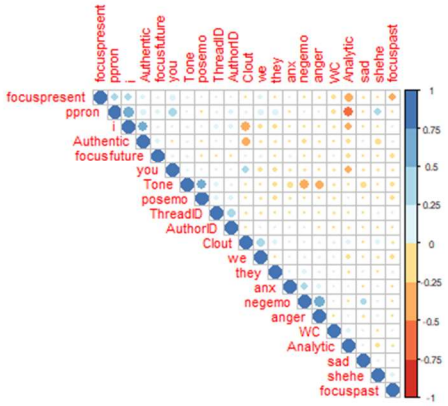
1: Perfect positive correlation. The variables tend to move in the same direction (i.e., when one variable increases, the other variable also increases).

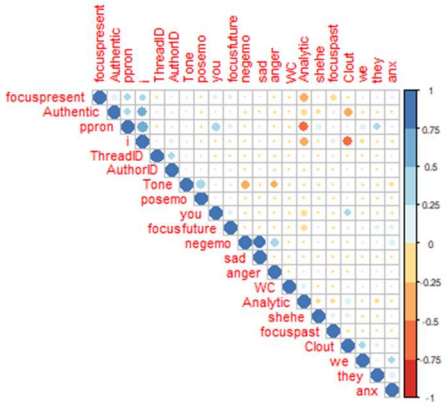
2002 **2003** **2004**

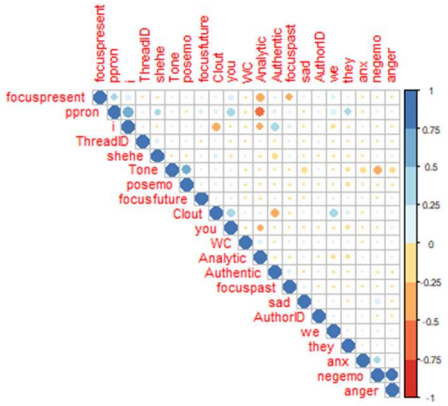
On 2002, we can see that the highest correlation from the variables is ppron – i, Tone – posemo and negemo – anger. Here shows that ppron – Analytic shows almost have perfect negative relation.

On 2003, we can see that the highest correlation is from Authentic – i, ppron – I and negemo – sad. Here shows that ppron – Analytic and i – Clout almost has perfect negative correlation.

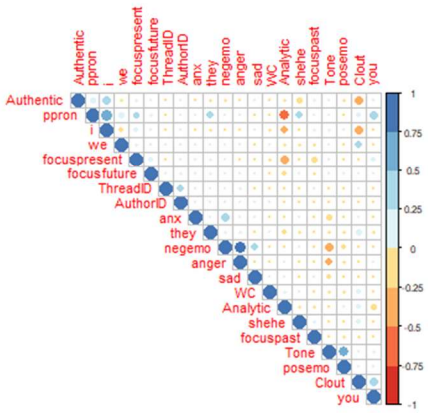
On 2004, we can see that the highest correlation is from ppron – I, Tone – posemo and negemo-anger. Here shows that ppron – Analytic almost has perfect negative correlation.



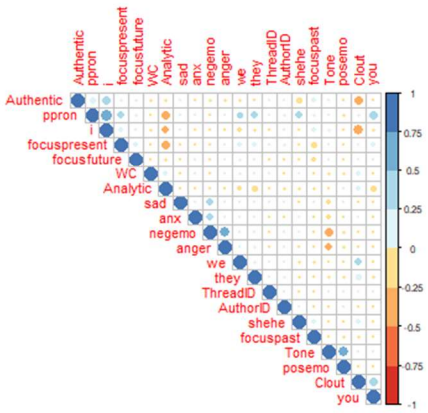




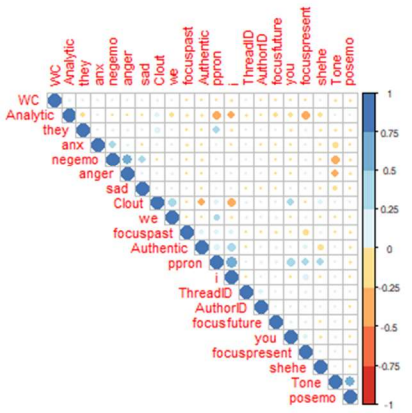
2005



2006



2007

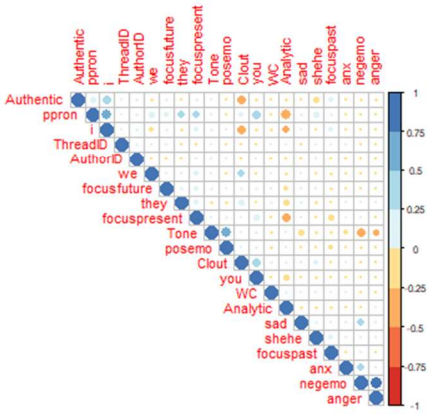


On 2005 we can see that the highest correlation shown by ppron – i, negemo – anger and Tone – posemo. Here shows that Analytic – ppron has the worse correlation almost has perfect negative correlation.

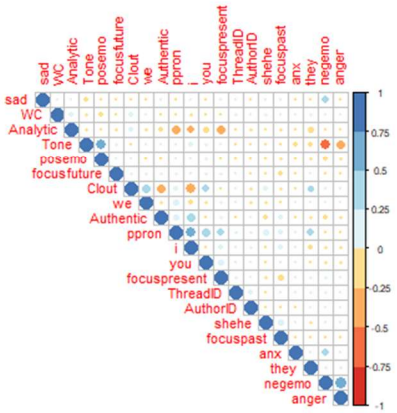
On 2006, we can see that the highest correlation shown by ppron – i, negemo – anger, Tone - posemo.

On 2007, we can see that the highest correlation shown by ppron – i and Tone – posemo.

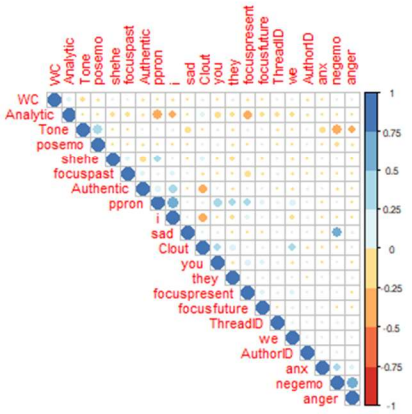
2008



2009



2010

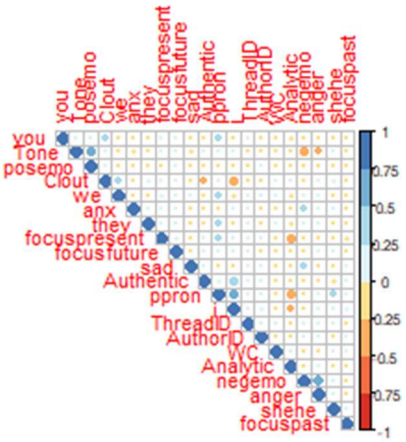


On 2008, we can that negemo – anger, ppron - i and Tone – posemo shows the highest correlations.

On 2009, we can see that Tone – posemo, ppron – I and negemo – anger show the most correlation. Tone – negemo shows almost the perfect negative correlation.

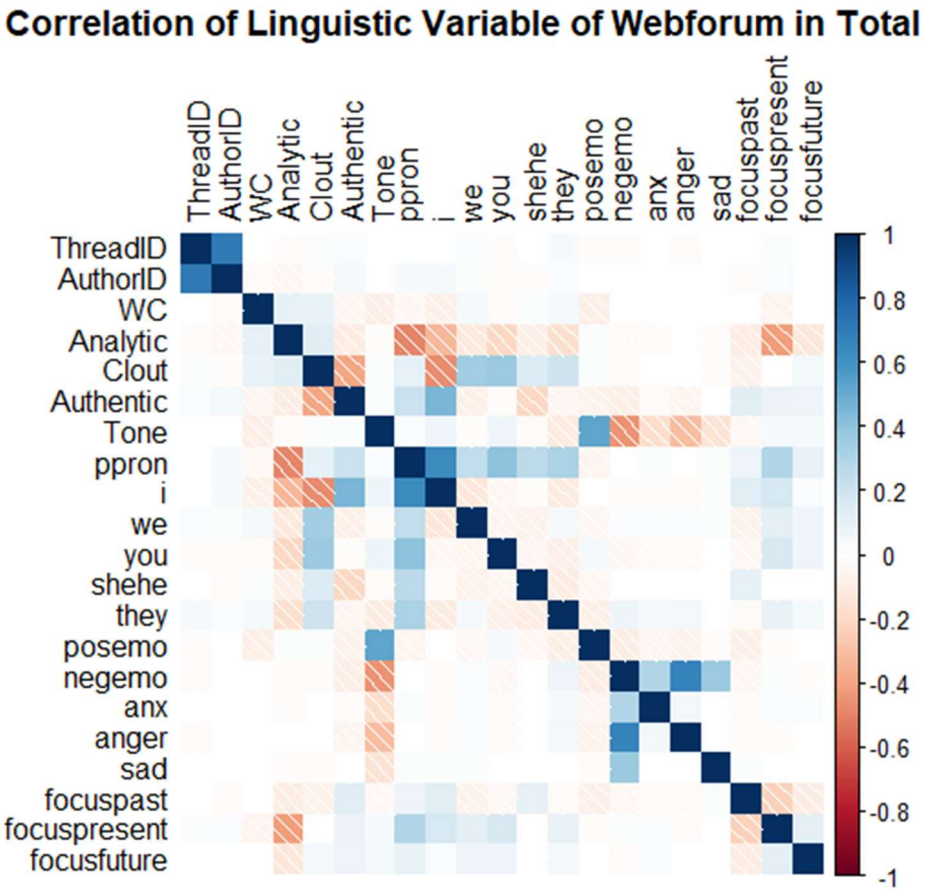
On 2010, we can see that ppron – i, sad – negemo and negemo – anger shows the most correlation.

2011



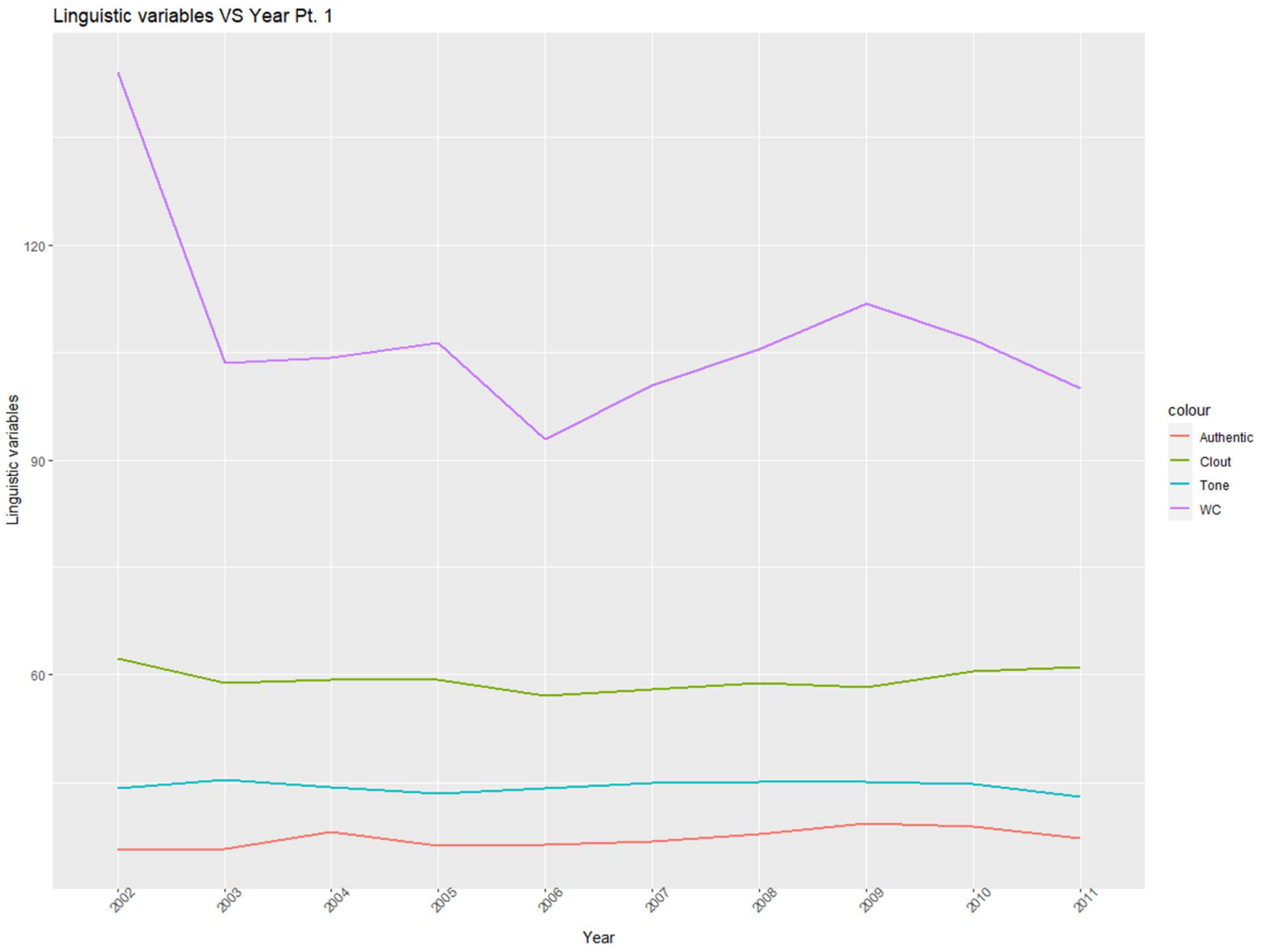
On 2011, we can see that ppron – i, Tone – posemo and negemo – anger shows the most correlation.

While we found out that there are some changes for the correlation every year which is for example shown by the multiple correlation graphs. We can now plot a different relation that shows the correlation for over the duration of forum.

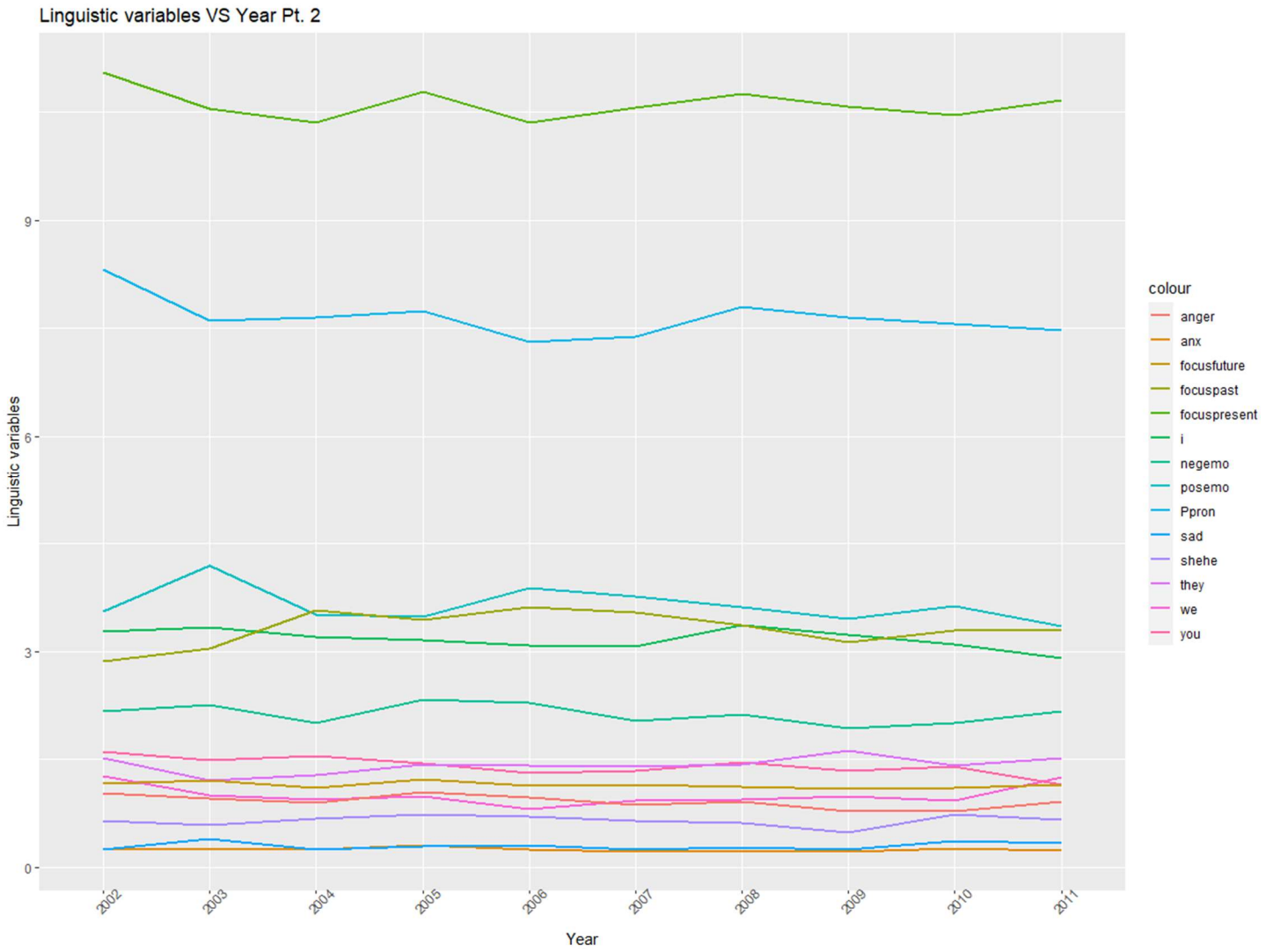


From the plot above we can see that excluding AuthorID and ThreadID. We can see that i-ppron, anger-negemo and posemo-Tone show the most correlations. However, those who makes almost close to perfect negative correlatives which ppron-Analystic, i-Clout, negemo-Tone, negemo-Tone, negemo-Tone, focuspresent – Analytic and anger - Tone.

Perfect negative correlatives which make it safe to say that those won't have any correlations and super weak relations. Now to make it clear, we can see the relation of the variable over the years more clearly with the graph below:



With the graph above we can see that there is not much difference of levels for Authentic, Clout and Tone. As if we see the data its shows that their values are doesn't change much in any spans most likely under 5 values difference. However, with the WC value variables we can see that there is a lot of drastic changes especially a drop from 2002 to 2006 (peak bottom) then start to rise again until 2009 and go down again to 2011.

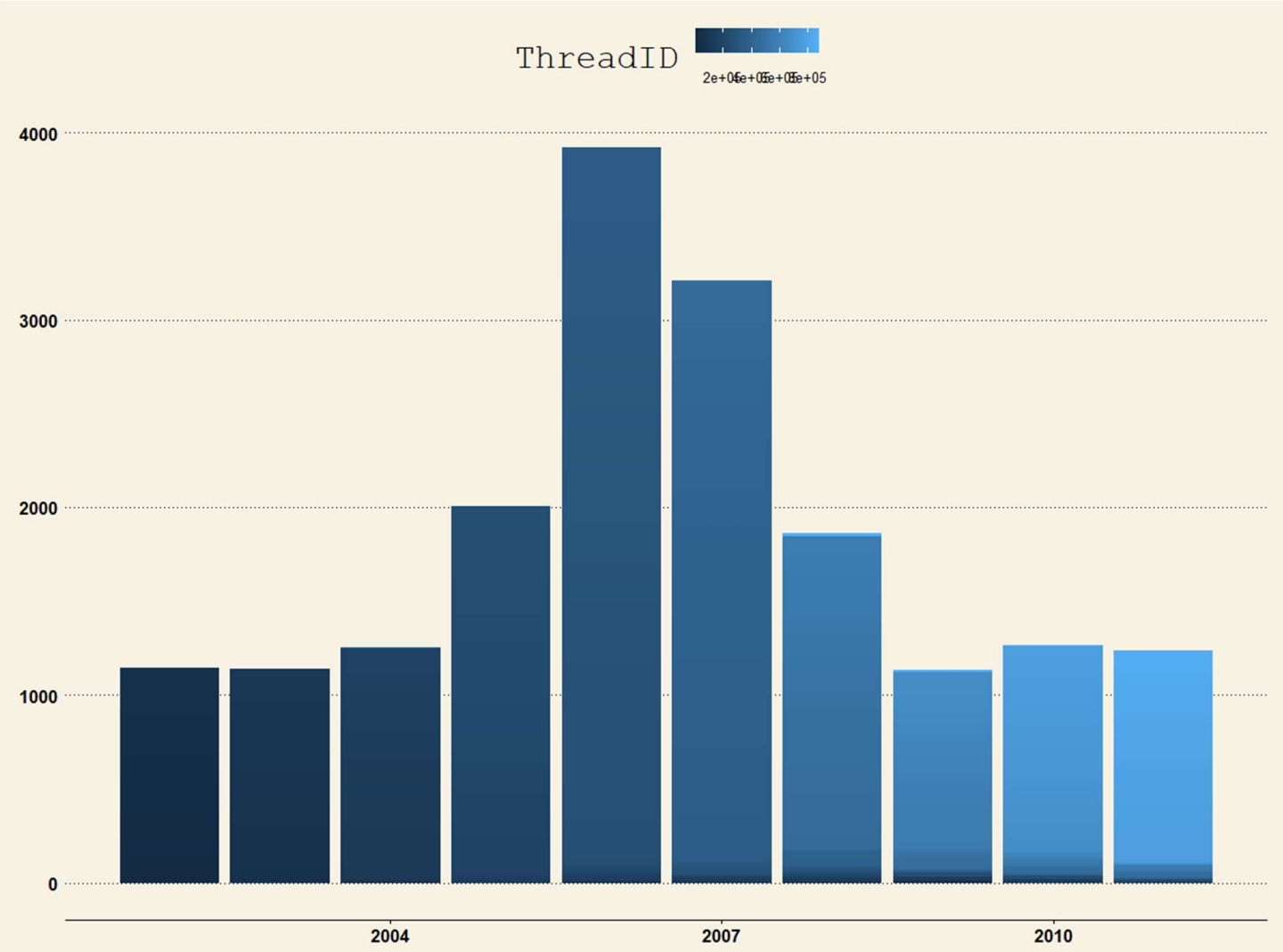


For the rest of the variables, there are no drastic changes unless for the 'Ppron' and 'posemo'. It shows a multiple ups and downs. Like how 'Ppron' drops significantly from 2002 and reach bottom on 2011. And how 'posemo' shows to highest peak in 2003 and 2006. 'Focuspresent' shows significant rise from 2002 to 2004 then has its lowest peak on 2009. 'they' also have lowest peak on 2009. Other than that, as the changes not that plenty it doesn't really shows any too much significant changes.

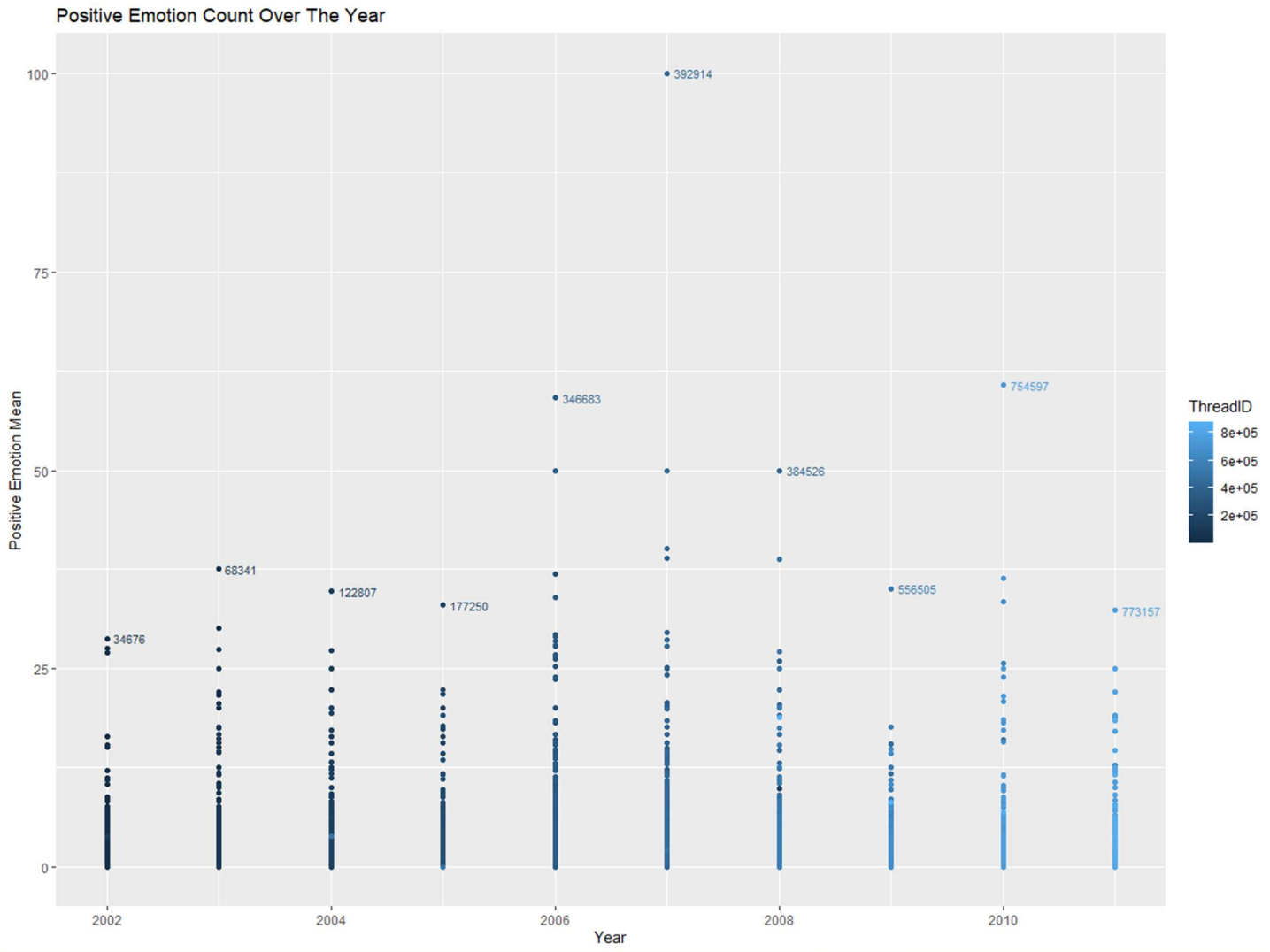
"In conclusion there is indeed some minor changes over the year for the level of the variables and the correlations, However this shows some correlation between those constant relation like with i-ppron, anger-negemo and posemo-Tone"

B.1. Using the relevant linguistic variables, is it possible to see whether or not particular threads are happier or more optimistic than other threads, or the forum in general, at different periods in time. (3 Marks)

In my definition of happier and more optimistic, from all of the variable I think it is safe to say that the ‘posemo’ is the most suitable variable of all as its expressing positive emotions rather than ‘negemo’, ‘anx’, ‘anger’, ‘sad’, ‘focuspast’ and ‘focusfuture’. Because most of them are on the negative side, I was considering to use ‘focusfuture’ as one of the indicators but I think we cannot define optimistic with just ‘focusfuture’ as it only shows how the data expressing focus in the future and doesn’t really answer the question or not of whether happier or optimistic. Therefore, I chose the most general options which is using the ‘posemo’ only.



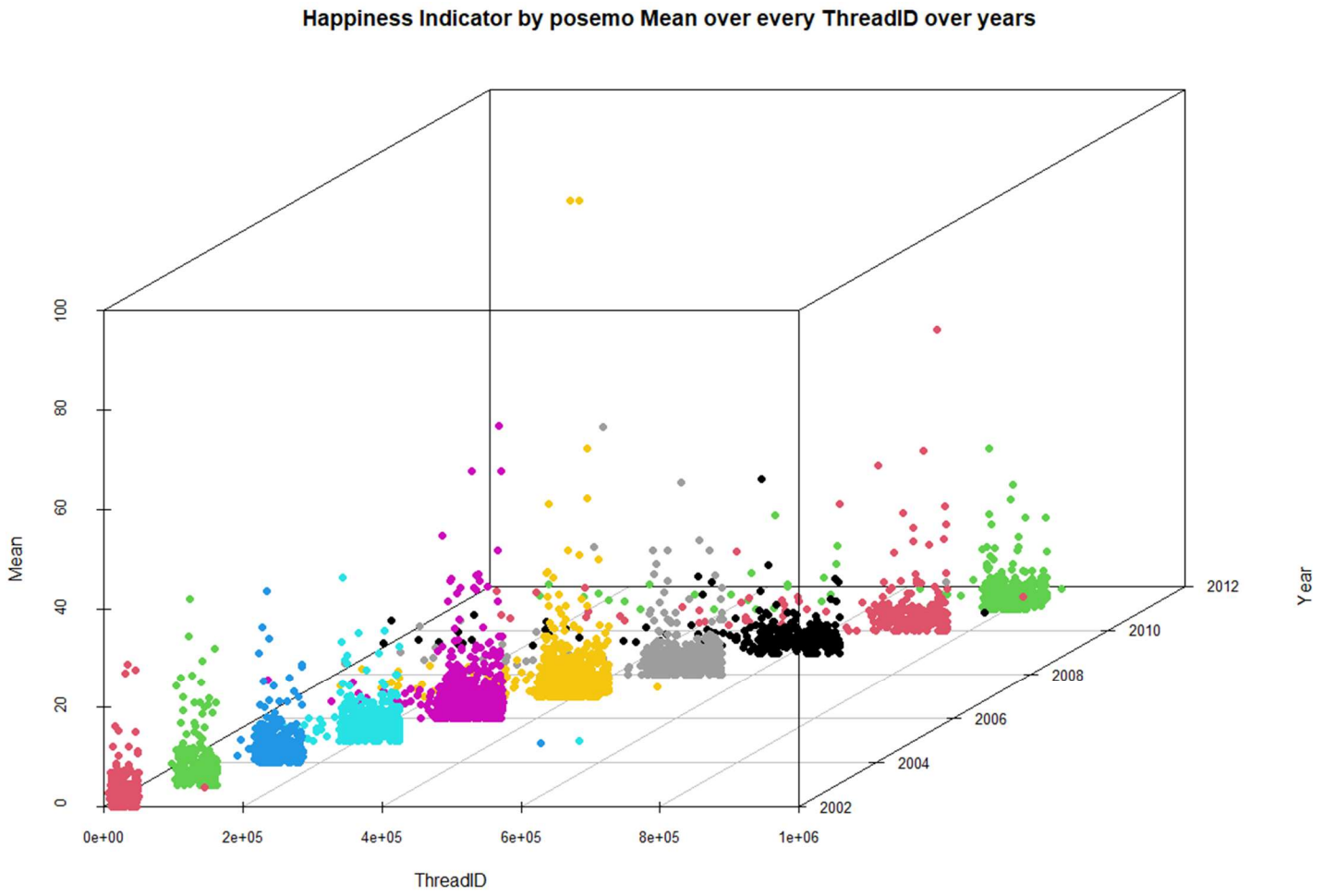
The graph above represents the highest mean of ‘posemo’ over the years of the web forum. The darker color shows higher Thread ID and vice – versa. However, we can’t really see the exact Thread ID as it shown as a bar. And so, we make another graph that shows only the points and the labels of the Thread ID however because there is too much stacked, I deduced it into a single highest Thread ID so we can see better. But from the graph above shown that the highest count for positive emotion is on 2006 in which has the IDs of higher so probably if the Thread ID generated from the lower to the higher it on the later part of the year. Below is the graph that shows the highest ‘posemo’ from each year.



With the indicator that the higher 'posemo' is means the happier the thread is we found some findings like below. This is a plot over the years with the highest Thread ID on every year. On 2002 is 34676, 2003 is 68431, 2004 is 122807, 2005 is 177250, 2006 is 346683, 2007 is 392914, 2008 is 384526, 2009 is 556505, 2010 is 754597 and 2011 is 773157. The highest Positive Emotion Mean shown in 2007. Relatively there probably is something good happen by that time and that Thread ID is probably the main discussion of it in this forum.

So yes, it is relatively possible to see whether or not particular Threads are happier than the other. However, due to the many Threads over the years which is around 4399 unique Threads ID. It is kind of hard to everything in one plot, however by dividing it into small pieces it definitely possible to see whether which Threads ID are happier and not just like how we can pin point the happiest Thread ID in every year.

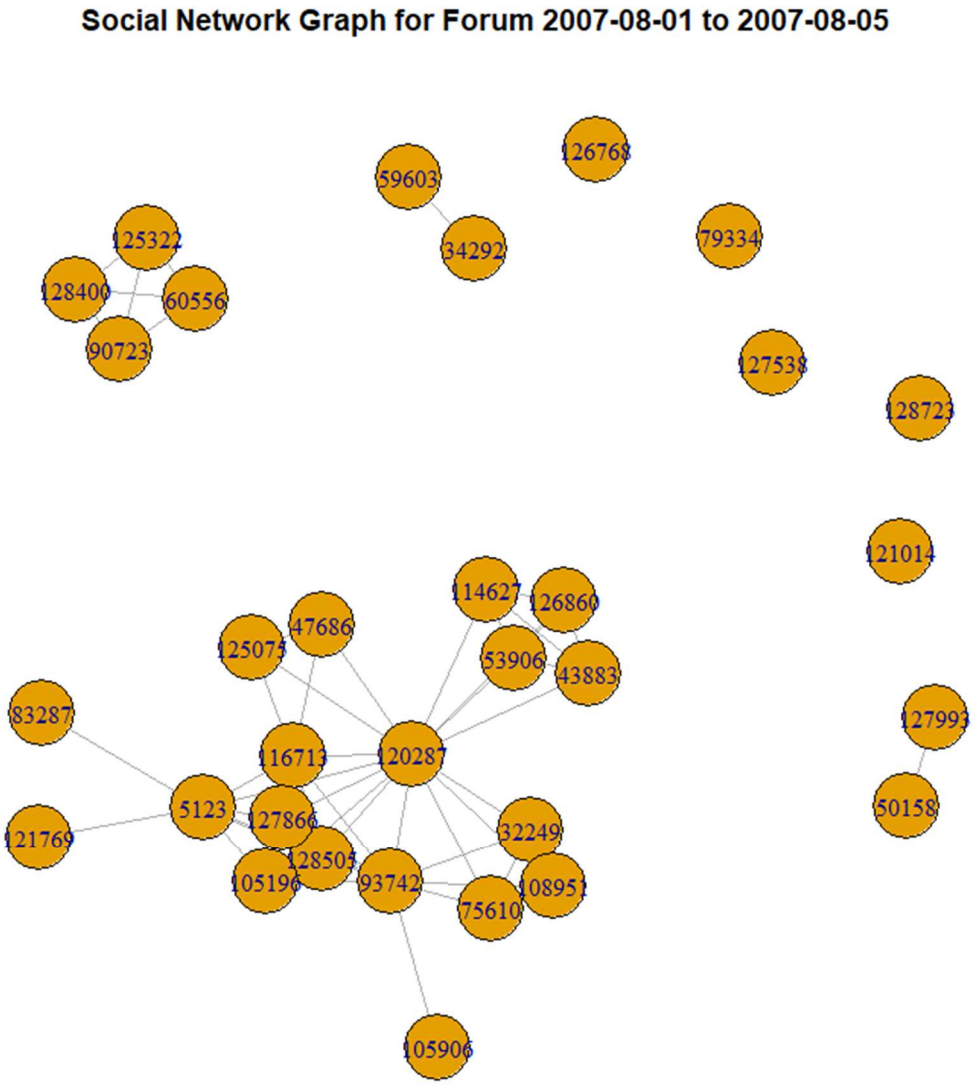
Below, we can see the scatter plots in 3D as an easier way to see the spreads of the Thread IDs by 'posemo' mean and year which is define by color also. And of course, in the graph below we can also see how 2007 has the highest record of 'posemo' mean which is shown by the yellow dots. And as we observe, we can also see that the older the Thread ID is the more likely it is to have bigger Thread ID number. We can also see that most of the 'posemo' usually have mean lower than 20.



It is possible to see whether some Threads happier than the other in some period of time.

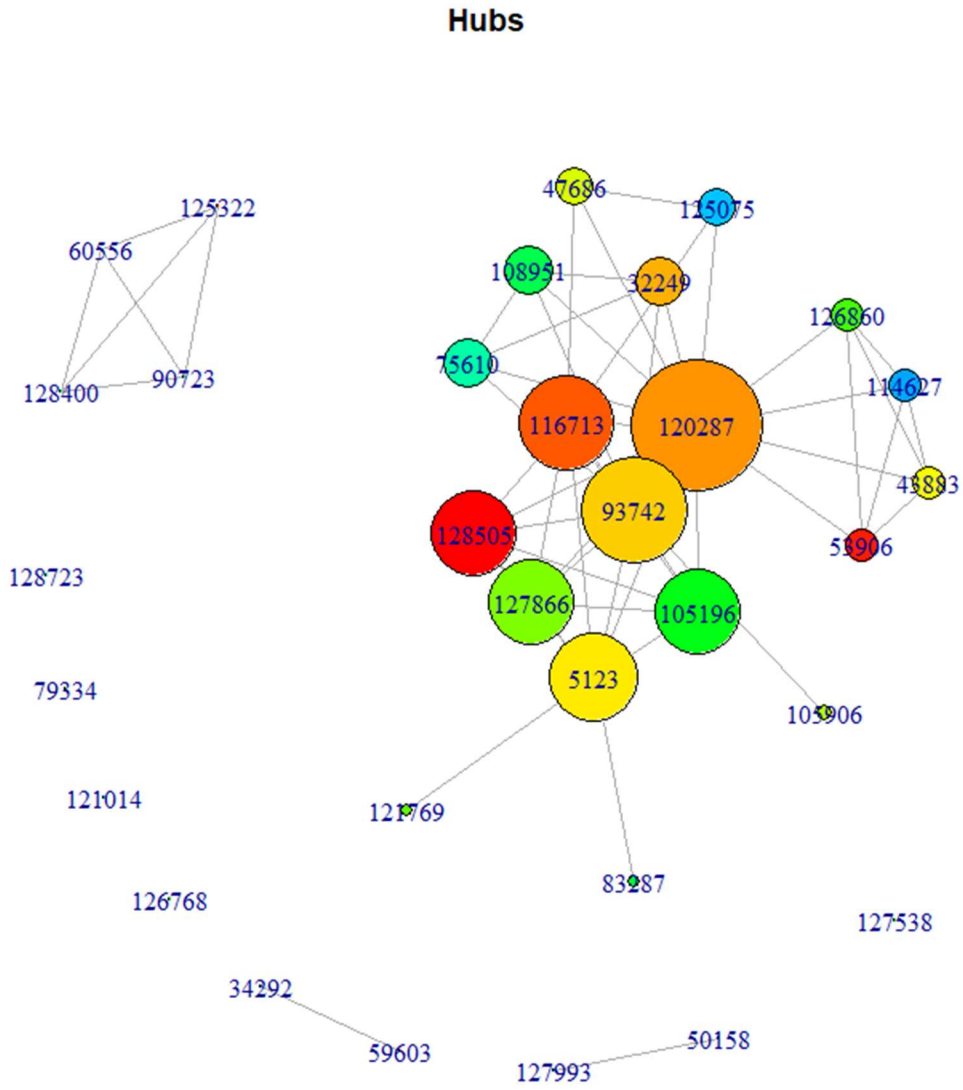
C.1. Create a non-trivial social network of all authors who are posting over a particular time period. For example, over one month. To create this, your social network should include at least 30 authors, some of whom will have posted to multiple (2 or more) threads during this period. Your social network should be connected, although some authors may be disconnected from the main group. Present your result as a network graph. (3 Marks)

Below is the Social Network for Forum on Threads from 2007-08-01 to 2007-08-05 with 32 nodes represents the Author ID have some post on some certain forums. The connection from the nodes is retrieve by author who post on the same Thread by that period of time.



C.2. Identify the most important author in the social network you created. Looking at the language they use, can you observe any difference between them and other members of their social network? (3 Marks)

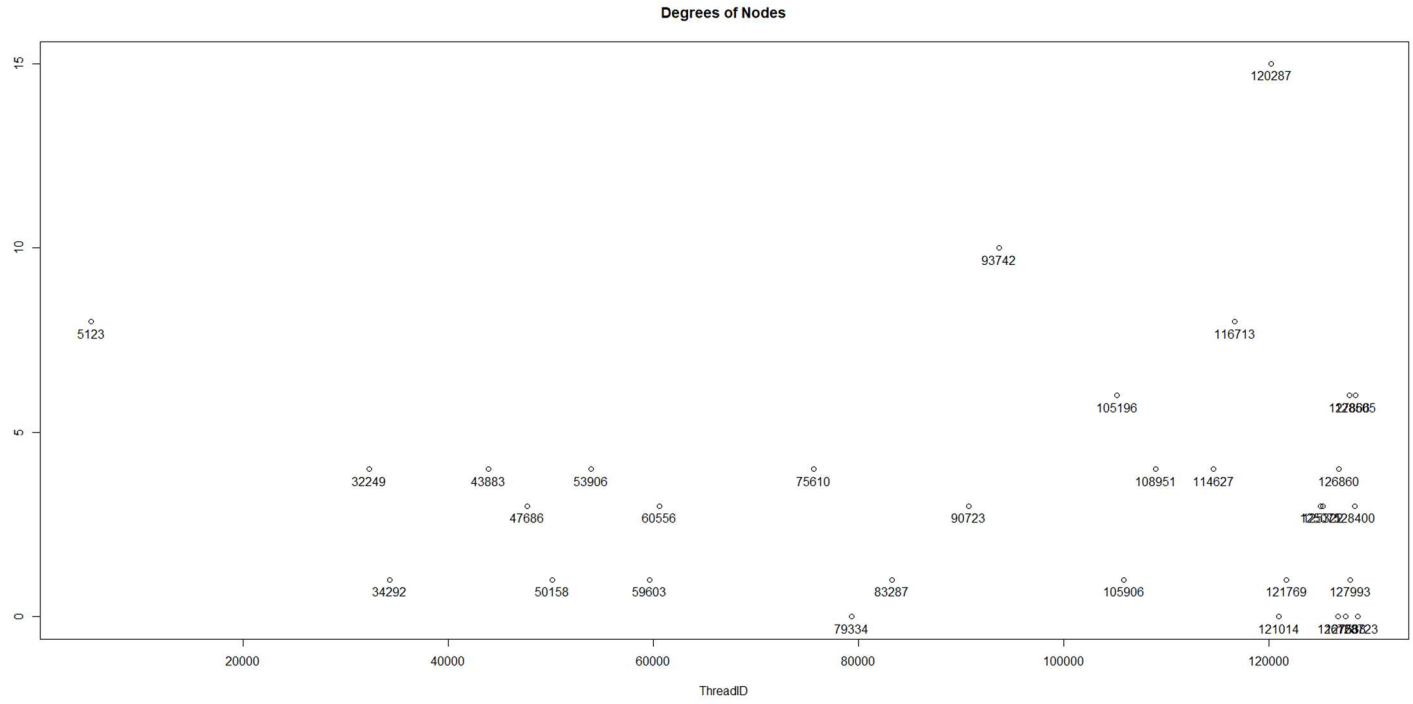
Below is the graph that have been modify by hubs. Basically, the hub has many outgoing links, and Authorities have many incoming links. From the modified plot below, we can see the most outgoing links are from 120287, 93742, 116713, 128505, 127866, 105196 and 5123. Making 120287 as the biggest/ most important author in the social network. Although way smaller and not as important some other group which are 75610, 108951, 47686, 125075, 32249, 126860, 114627, 43883 and 53906. And the rest compared too significantly smaller than the rest.



A node's degree is simply a count of how many social connections (i.e., edges) it has. The degree centrality for a node is simply its degree. A node with 10 social connections would have a degree centrality of 10. A node with 1 edge would have a degree centrality of 1. It is one of the indicators that can show how much important a node in a graph is. Because in this forum case. It means that the degree has activities in the same forum with another nodes. And generally, it can be said that it means this Author is quite active in the group.

It is used for finding very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

Degree centrality is the simplest measure of node connectivity. Sometimes it's useful to look at in-degree (number of inbound links) and out-degree (number of outbound links) as distinct measures, for example when looking at transactional data or account activity.



From the table above, we can see that 120287 has the highest degree which is 15, followed by 93742 which as 10 degrees and then followed by 116713 and 5123 which both have 8 degrees. If we judged Author importance by its degree than it means that 120287 is the most important author followed by those three.

```
> degree
128505  53906 125322 116713  59603 120287  32249  93742   5123  43883  47686 105906  79334 127866 121769 126860 127538 126768 105196
     6      4      3      8      1     15      4     10      8      4      3      1      0      6      1      4      0      0      6
 83287 108951  50158 128400  75610 128723  90723  60556 121014 125075 114627 127993  34292
     1      4      1      3      4      0      3      3      0      3      4      1      1
```

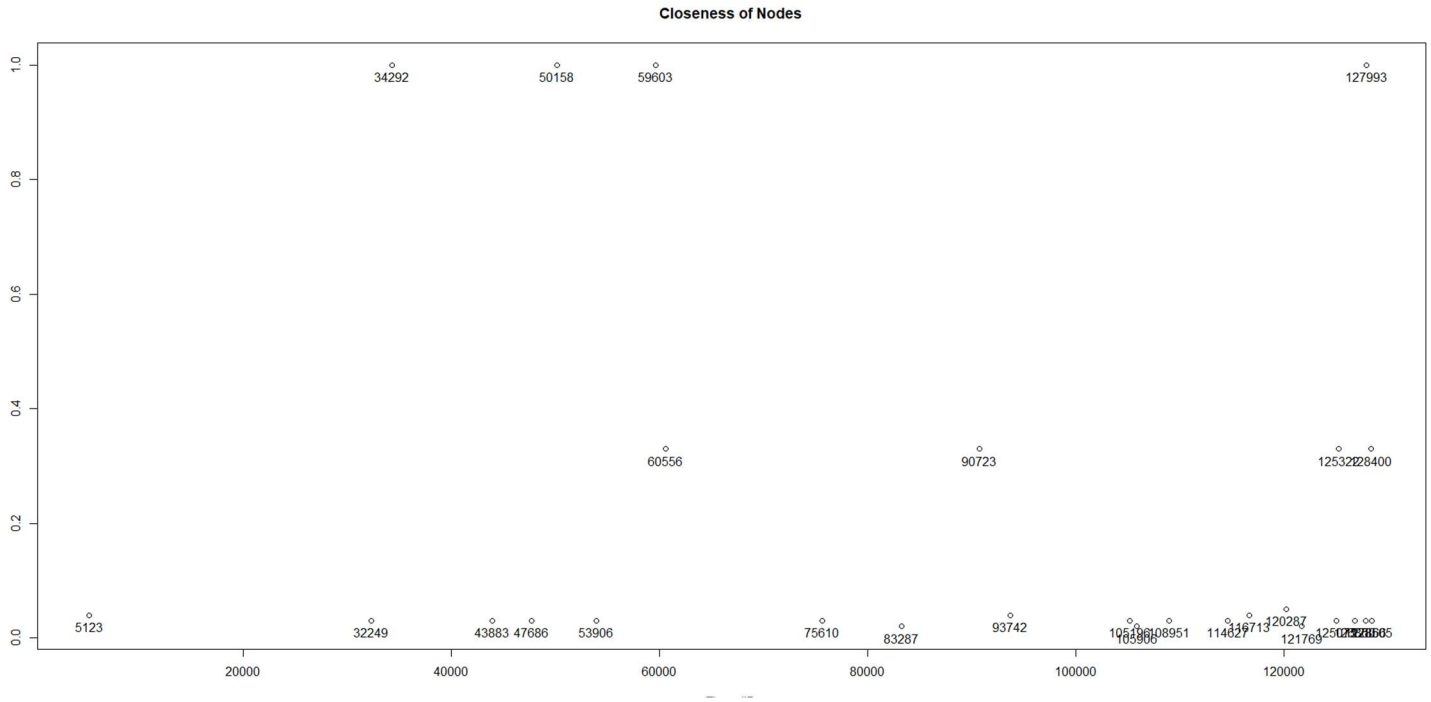
```
> filter(social_data,social_data$AuthorID==120287)
   ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe  they posemo negemo  anx anger  sad focuspast focuspresent focusfuture Year
1    404456   120287 2007-08-05 03:18  14    13.85  1.67     99.00 99.00 14.29 14.29 0.00 0.00     0  0.00   7.14   0.00 0.00  0.00 0.00      7.14         7.14        0.00 2007
2    404456   120287 2007-08-05 03:01  17     1.00 11.97     28.81 98.87 11.76 11.76 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00        17.65        0.00 2007
3    399271   120287 2007-08-03 11:21  49    92.84 96.66      7.24 91.78  2.04  0.00 0.00 2.04     0  0.00   6.12   2.04 0.00  0.00 0.00      2.04         8.16        0.00 2007
4    408419   120287 2007-08-03 01:22 727    88.27 54.92      7.55 76.59  2.34  0.55 0.55 0.69     0  0.55   5.23   2.48 0.14  1.38 0.41      1.24         6.88        1.93 2007
5    408224   120287 2007-08-02 05:04  11    99.00 50.00     85.21 25.77  0.00  0.00 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      0.00        18.18        0.00 2007
6    408419   120287 2007-08-03 01:24 116    94.36 56.86     11.50 25.77  1.72  0.86 0.86 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      2.59         4.31        3.45 2007
7    408491   120287 2007-08-03 10:49  14    83.00 23.75     52.86 25.77  7.14  7.14 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      7.14         0.00        0.00 2007
8    404456   120287 2007-08-03 02:52  10    98.46 15.86      3.37  1.00 10.00 10.00 0.00 0.00     0  0.00   0.00  20.00 0.00 20.00 0.00      0.00         0.00        0.00 2007
9    404456   120287 2007-08-03 02:46   4     1.80 99.00      1.00 25.77 25.00  0.00 0.00 0.00     0 25.00   0.00   0.00 0.00  0.00 0.00      0.00         0.00       25.00 2007
10   408997   120287 2007-08-05 11:18  17    97.02 50.00     66.34 98.87  0.00  0.00 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00         5.88        0.00 2007
> filter(social_data,social_data$AuthorID==93742)
  ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe they posemo negemo  anx anger sad focuspast focuspresent focusfuture Year
1   408419    93742 2007-08-03 04:15  27    65.78 13.32     79.42 25.77  7.41  0.00 0.00 0.00     0 7.41   0.00   0.00 0.00  0.00   0      0.00        14.81        0.00 2007
2   408419    93742 2007-08-03 06:36  70    77.33 66.59     18.92  8.64  5.71  1.43 2.86 0.00     0 1.43   0.00   1.43 0.00  1.43   0      1.43         7.14        0.00 2007
3   408491    93742 2007-08-04 07:57   9    17.96 13.32      4.97  1.00 11.11 11.11 0.00 0.00     0 0.00   0.00  11.11 0.00  0.00   0      0.00        22.22       11.11 2007
4   408491    93742 2007-08-03 06:59 187    17.30 54.27     52.60 66.27 13.90  8.02 0.00 5.35     0 0.53   2.67   0.53 0.53  0.00   0      1.07        17.11        1.07 2007
5   336155    93742 2007-08-03 04:23  27    55.62 96.78     92.47 88.52 18.52  7.41 0.00 7.41     0 3.70   3.70   0.00 0.00  0.00   0      3.70        14.81        0.00 2007
6   408419    93742 2007-08-03 03:02 123    77.86 65.80     21.47 56.48  4.88  1.63 0.00 3.25     0 0.00   3.25   1.63 0.00  0.81   0      0.81         7.32        1.63 2007
> filter(social_data,social_data$AuthorID==116713)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic Tone ppron     i we  you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   404456   116713 2007-08-03 19:20 42    73.03  59.4     37.24    1  9.52  7.14  0  0.0     0 2.38   2.38   7.14   0     0   0      7.14        14.29           0 2007
2   408419   116713 2007-08-04 20:43  8    71.61  99.0      7.84   99 25.00 12.50  0 12.5     0 0.00  12.50   0.00   0     0   0     12.50         0.00           0 2007
> filter(social_data,social_data$AuthorID==5123)
  ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe they posemo negemo anx anger  sad focuspast focuspresent focusfuture Year
1   408264      5123 2007-08-04 02:51  94    68.30 10.10     71.83 99.00  3.19 2.13 0.00 1.06     0 0.00   7.45   1.06   0  0.00 0.00      0.00        11.70        1.06 2007
2   408419      5123 2007-08-04 03:14 922    76.44 75.63     11.78 35.23  5.21 0.65 0.22 2.82     0 1.52   2.49   1.95   0  1.08 0.11      0.65         9.22        1.19 2007
3   407407      5123 2007-08-04 02:28 515    66.93 76.97     15.66 72.97  6.41 1.75 0.58 0.39     0 3.69   3.11   0.58   0  0.19 0.19      1.17        10.10        0.39 2007
```

As we can see the difference between this node and another node and why is that they have a lot of degree is that they are either active in many threads like 120287 or be in one of the active forum like 116713, it only attend 2 threads yet it has many degree because that Threads is popular hence connects it with a lot of people.

Closeness centrality scores each node based on their ‘closeness’ to all other nodes in the network. This measure calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths.

It is for finding the individuals who are best placed to influence the entire network most quickly. Closeness centrality can help find good ‘broadcasters’, but in a highly-connected network, you will often find all nodes have a similar score. What may be more useful is using Closeness to find influencers in a single cluster.

```
> closeness
128505  53906 125322 116713  59603 120287  32249  93742   5123  43883  47686 105906  79334 127866 121769 126860 127538 126768 105196
  0.03   0.03   0.33   0.04   1.00   0.05   0.03   0.04   0.04   0.03   0.03   0.02    NaN   0.03   0.02   0.03    NaN    NaN   0.03
 83287 108951  50158 128400  75610 128723  90723  60556 121014 125075 114627 127993  34292
  0.02   0.03   1.00   0.33   0.03    NaN   0.33   0.33    NaN   0.03   0.03   1.00   1.00
```



From this graph we can see that the highest nodes that has highest Closeness score are 127993, 34292, 50158 and 59603.

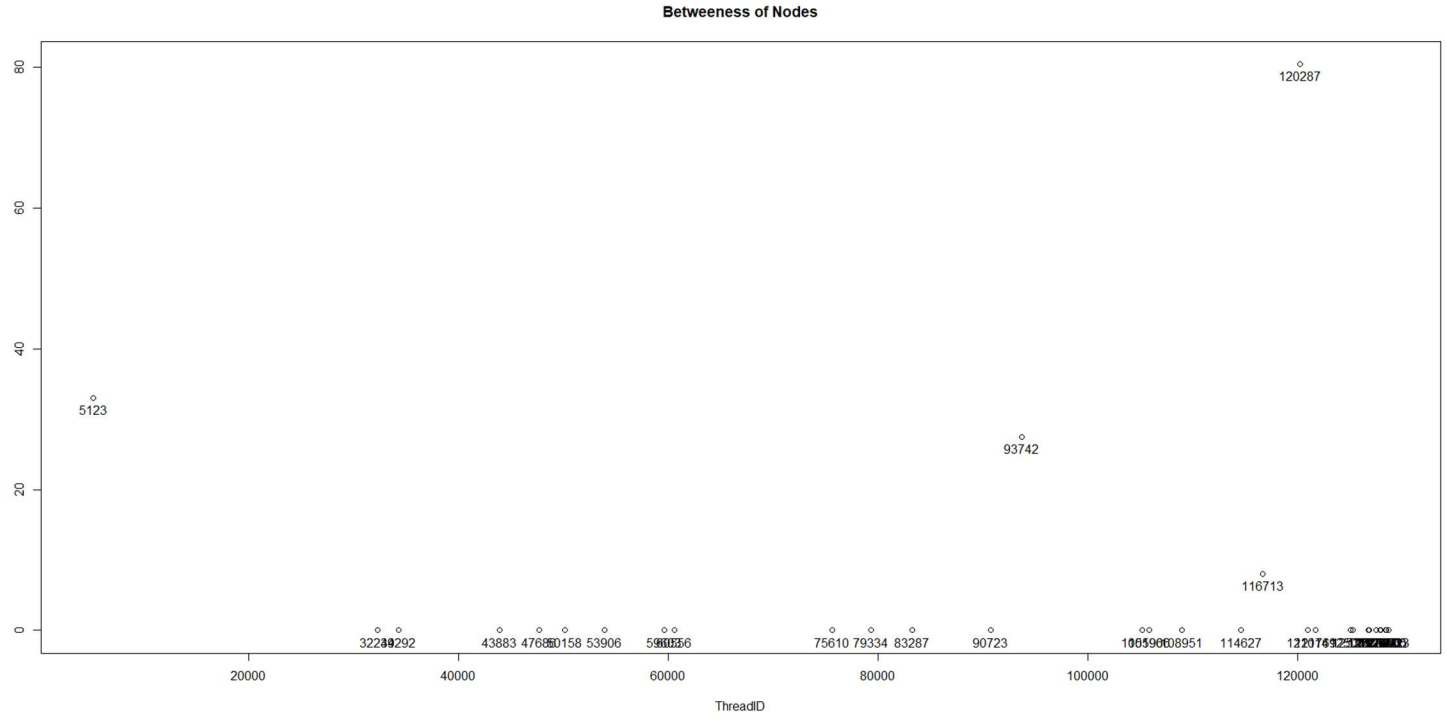
```
> filter(social_data,social_data$AuthorID==127993)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic  Tone ppron    i   we  you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   407427   127993 2007-08-02 17:25 44     25.9  94.4     61.34 68.66 20.45 9.09 4.55 2.27  2.27 2.27   2.27      0   0     0   0      6.82         9.09          0 2007
> filter(social_data,social_data$AuthorID==34292)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic  Tone ppron i we you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   408490    34292 2007-08-03 12:23  4    37.39     1        99 25.77     0 0  0   0     0   0      0      0   0     0   0         0           25           0 2007
> filter(social_data,social_data$AuthorID==50158)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic  Tone ppron   i we you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   407427    50158 2007-08-02 11:01 27    92.84 64.46     58.07 25.77  7.41 3.7  0 3.7     0   0      0      0   0     0   0       3.7          3.7          0 2007
> filter(social_data,social_data$AuthorID==59603)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic  Tone ppron     i we  you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   408490    59603 2007-08-03 06:47 60    32.98 30.86     99.00 84.54 13.33 11.67  0 1.67     0   0   3.33   0.00   0     0   0      5.00        10.00         1.67 2007
2   408490    59603 2007-08-03 07:29 35    28.22 38.74     95.88  1.91  8.57  8.57  0 0.00     0   0   0.00   2.86   0     0   0      2.86         8.57         2.86 2007
>
```

As we can see above, it is kind of hard to see what make these have the highest degree as it was really varies.

However, as closeness shows that these nodes are closest to the other and based on what I observed from the table `above. We can see that the 4 of them have nodes that has similar Dates and Times and approximately in the middle thus making them the closest to the others

Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.This measure shows which nodes are ‘bridges’ between nodes in a network. It does this by identifying all the shortest paths and then counting how many times each node falls on one. It is for finding the individuals who influence the flow around a system. Betweenness is useful for analyzing communication dynamics, but should be used with care. A high betweenness count could indicate someone holds authority over disparate clusters in a network, or just that they are on the periphery of both clusters.

```
> betweenness
128505  53906 125322 116713  59603 120287  32249  93742   5123  43883  47686 105906  79334 127866 121769 126860 127538 126768 105196
   0.0     0.0     0.0     8.0     0.0    80.5     0.0    27.5    33.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
 83287 108951  50158 128400  75610 128723  90723  60556 121014 125075 114627 127993  34292
   0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0     0.0
```



We can see that 120287 has the highest betweenness score followed by 5123 and 93742.

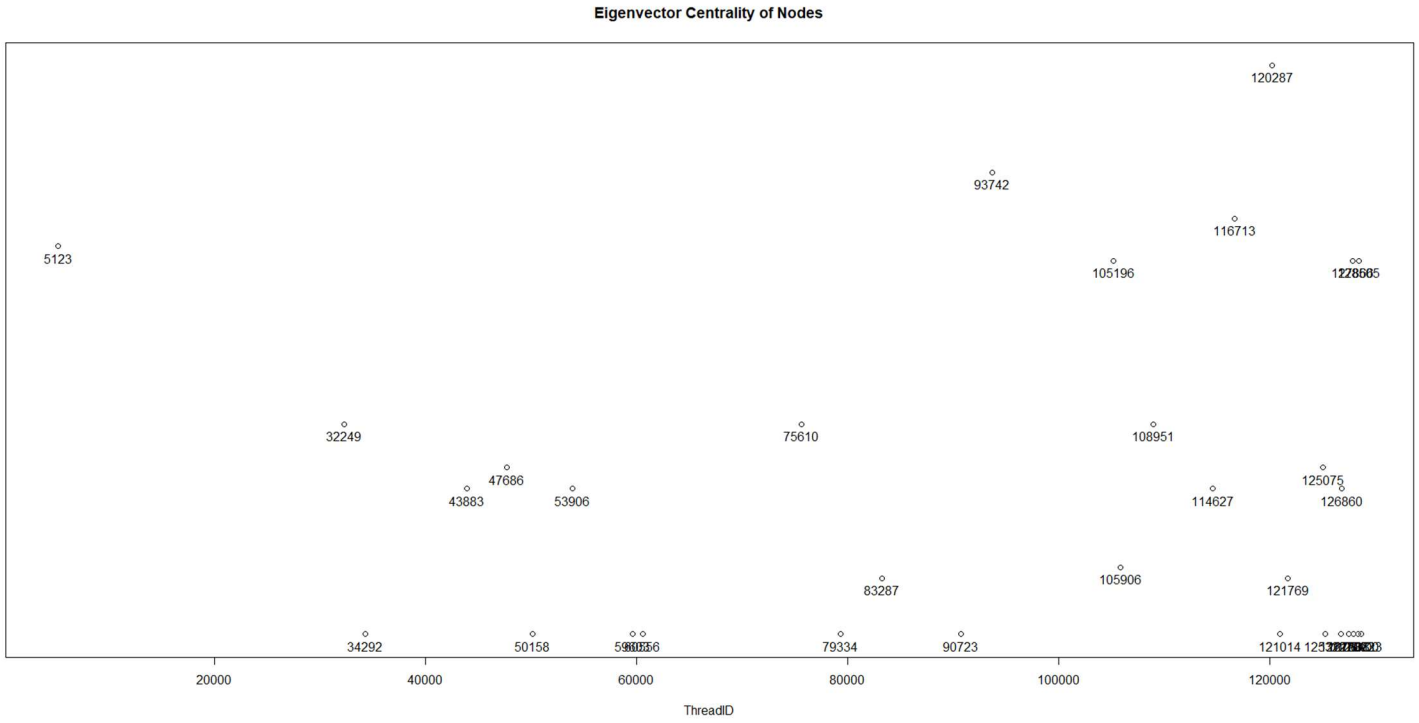
```
> filter(social_data,social_data$AuthorID==120287)
   ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe  they posemo negemo  anx anger  sad focuspast focuspresent focusfuture Year
1    404456   120287 2007-08-05 03:18  14    13.85  1.67     99.00 99.00 14.29 14.29 0.00 0.00     0  0.00   7.14   0.00 0.00  0.00 0.00      7.14         7.14       0.00 2007
2    404456   120287 2007-08-05 03:01  17     1.00 11.97     28.81 98.87 11.76 11.76 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00        17.65       0.00 2007
3    399271   120287 2007-08-03 11:21  49    92.84 96.66      7.24 91.78  2.04  0.00 0.00 2.04     0  0.00   6.12   2.04 0.00  0.00 0.00      2.04         8.16       0.00 2007
4    408419   120287 2007-08-03 01:22 727    88.27 54.92      7.55 76.59  2.34  0.55 0.55 0.69     0  0.55   5.23   2.48 0.14  1.38 0.41      1.24         6.88       1.93 2007
5    408224   120287 2007-08-02 05:04  11    99.00 50.00     85.21 25.77  0.00  0.00 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      0.00        18.18       0.00 2007
6    408419   120287 2007-08-03 01:24 116    94.36 56.86     11.50 25.77  1.72  0.86 0.86 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      2.59         4.31       3.45 2007
7    408491   120287 2007-08-03 10:49  14    83.00 23.75     52.86 25.77  7.14  7.14 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      7.14         0.00       0.00 2007
8    404456   120287 2007-08-03 02:52  10    98.46 15.86      3.37  1.00 10.00 10.00 0.00 0.00     0  0.00   0.00  20.00 0.00 20.00 0.00      0.00         0.00       0.00 2007
9    404456   120287 2007-08-03 02:46   4     1.80 99.00      1.00 25.77 25.00  0.00 0.00 0.00     0 25.00   0.00   0.00 0.00  0.00 0.00      0.00         0.00      25.00 2007
10   408997   120287 2007-08-05 11:18  17    97.02 50.00     66.34 98.87  0.00  0.00 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00         5.88       0.00 2007
> filter(social_data,social_data$AuthorID==5123)
  ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron    i   we  you shehe they posemo negemo anx anger  sad focuspast focuspresent focusfuture Year
1   408264      5123 2007-08-04 02:51  94    68.30 10.10     71.83 99.00  3.19 2.13 0.00 1.06     0 0.00   7.45   1.06   0  0.00 0.00      0.00        11.70       1.06 2007
2   408419      5123 2007-08-04 03:14 922    76.44 75.63     11.78 35.23  5.21 0.65 0.22 2.82     0 1.52   2.49   1.95   0  1.08 0.11      0.65         9.22       1.19 2007
3   407407      5123 2007-08-04 02:28 515    66.93 76.97     15.66 72.97  6.41 1.75 0.58 0.39     0 3.69   3.11   0.58   0  0.19 0.19      1.17        10.10       0.39 2007
> filter(social_data,social_data$AuthorID==93742)
  ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron    i   we  you shehe they posemo negemo  anx anger sad focuspast focuspresent focusfuture Year
1   408419     93742 2007-08-03 04:15  27    65.78 13.32     79.42 25.77  7.41  0.00 0.00 0.00     0 7.41   0.00   0.00 0.00  0.00   0      0.00        14.81       0.00 2007
2   408419     93742 2007-08-03 06:36  70    77.33 66.59     18.92  8.64  5.71  1.43 2.86 0.00     0 1.43   0.00   1.43 0.00  1.43   0      1.43         7.14       0.00 2007
3   408491     93742 2007-08-04 07:57   9    17.96 13.32      4.97  1.00 11.11 11.11 0.00 0.00     0 0.00   0.00  11.11 0.00  0.00   0      0.00        22.22      11.11 2007
4   408491     93742 2007-08-03 06:59 187    17.30 54.27     52.60 66.27 13.90  8.02 0.00 5.35     0 0.53   2.67   0.53 0.53  0.00   0      1.07        17.11       1.07 2007
5   336155     93742 2007-08-03 04:23  27    55.62 96.78     92.47 88.52 18.52  7.41 0.00 7.41     0 3.70   3.70   0.00 0.00  0.00   0      3.70        14.81       0.00 2007
6   408419     93742 2007-08-03 03:02 123    77.86 65.80     21.47 56.48  4.88  1.63 0.00 3.25     0 0.00   3.25   1.63 0.00  0.81   0      0.81         7.32       1.63 2007
```

As you can see above, those 3 are some of the highest degree nodes we found on the form. And the similarities they all have is that they were in the same Thread which is 408109. Other than that, it was really varies.

Like degree centrality, EigenCentrality measures a node's influence based on the number of links it has to other nodes in the network. EigenCentrality then goes a step further by also taking into account how well connected a node is, and how many links their connections have, and so on through the network. By calculating the extended connections of a node, EigenCentrality can identify nodes with influence over the whole network, not just those directly connected to it.

EigenCentrality is a good 'all-round' SNA score, handy for understanding human social networks, but also for understanding networks like malware propagation.

```
> eig
      128505       53906      125322      116713       59603      120287       32249       93742        5123       43883
6.566352e-01 2.557921e-01 6.096961e-17 7.306911e-01 0.000000e+00 1.000000e+00 3.689702e-01 8.114314e-01 6.815788e-01 2.557921e-01
       47686      105906       79334      127866      121769      126860      127538      126768      105196       83287
2.928696e-01 1.174383e-01 0.000000e+00 6.566352e-01 9.864479e-02 2.557921e-01 0.000000e+00 0.000000e+00 6.566352e-01 9.864479e-02
      108951       50158      128400       75610      128723       90723       60556      121014      125075      114627
3.689702e-01 1.524240e-17 6.096961e-17 3.689702e-01 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 2.928696e-01 2.557921e-01
      127993       34292
0.000000e+00 1.524240e-17
```



With these table we can see that 120287, 93742, 5123 and 116713 have the highest eig.

```
> filter(social_data,social_data$AuthorID==120287)
   ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe  they posemo negemo  anx anger  sad focuspast focuspresent focusfuture Year
1    404456   120287 2007-08-05 03:18  14    13.85  1.67     99.00 99.00 14.29 14.29 0.00 0.00     0  0.00   7.14   0.00 0.00  0.00 0.00      7.14          7.14         0.00 2007
2    404456   120287 2007-08-05 03:01  17     1.00 11.97     28.81 98.87 11.76 11.76 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00         17.65         0.00 2007
3    399271   120287 2007-08-03 11:21  49    92.84 96.66      7.24 91.78  2.04  0.00 0.00 2.04     0  0.00   6.12   2.04 0.00  0.00 0.00      2.04          8.16         0.00 2007
4    408419   120287 2007-08-03 01:22 727    88.27 54.92      7.55 76.59  2.34  0.55 0.55 0.69     0  0.55   5.23   2.48 0.14  1.38 0.41      1.24          6.88         1.93 2007
5    408224   120287 2007-08-02 05:04  11    99.00 50.00     85.21 25.77  0.00  0.00 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      0.00         18.18         0.00 2007
6    408419   120287 2007-08-03 01:24 116    94.36 56.86     11.50 25.77  1.72  0.86 0.86 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      2.59          4.31         3.45 2007
7    408491   120287 2007-08-03 10:49  14    83.00 23.75     52.86 25.77  7.14  7.14 0.00 0.00     0  0.00   0.00   0.00 0.00  0.00 0.00      7.14          0.00         0.00 2007
8    404456   120287 2007-08-03 02:52  10    98.46 15.86      3.37  1.00 10.00 10.00 0.00 0.00     0  0.00   0.00  20.00 0.00 20.00 0.00      0.00          0.00         0.00 2007
9    404456   120287 2007-08-03 02:46   4     1.80 99.00      1.00 25.77 25.00  0.00 0.00 0.00     0 25.00   0.00   0.00 0.00  0.00 0.00      0.00          0.00        25.00 2007
10   408997   120287 2007-08-05 11:18  17    97.02 50.00     66.34 98.87  0.00  0.00 0.00 0.00     0  0.00   5.88   0.00 0.00  0.00 0.00      0.00          5.88         0.00 2007
> filter(social_data,social_data$AuthorID==93742)
   ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron     i   we  you shehe they posemo negemo  anx anger sad focuspast focuspresent focusfuture Year
1    408419    93742 2007-08-03 04:15  27    65.78 13.32     79.42 25.77  7.41  0.00 0.00 0.00     0 7.41   0.00   0.00 0.00  0.00   0      0.00         14.81         0.00 2007
2    408419    93742 2007-08-03 06:36  70    77.33 66.59     18.92  8.64  5.71  1.43 2.86 0.00     0 1.43   0.00   1.43 0.00  1.43   0      1.43          7.14         0.00 2007
3    408491    93742 2007-08-04 07:57   9    17.96 13.32      4.97  1.00 11.11 11.11 0.00 0.00     0 0.00   0.00  11.11 0.00  0.00   0      0.00         22.22        11.11 2007
4    408491    93742 2007-08-03 06:59 187    17.30 54.27     52.60 66.27 13.90  8.02 0.00 5.35     0 0.53   2.67   0.53 0.53  0.00   0      1.07         17.11         1.07 2007
5    336155    93742 2007-08-03 04:23  27    55.62 96.78     92.47 88.52 18.52  7.41 0.00 7.41     0 3.70   3.70   0.00 0.00  0.00   0      3.70         14.81         0.00 2007
6    408419    93742 2007-08-03 03:02 123    77.86 65.80     21.47 56.48  4.88  1.63 0.00 3.25     0 0.00   3.25   1.63 0.00  0.81   0      0.81          7.32         1.63 2007
> filter(social_data,social_data$AuthorID==5123)
  ThreadID AuthorID       Date  Time  WC Analytic Clout Authentic  Tone ppron    i   we  you shehe they posemo negemo anx anger  sad focuspast focuspresent focusfuture Year
1   408264     5123 2007-08-04 02:51  94    68.30 10.10     71.83 99.00  3.19 2.13 0.00 1.06     0 0.00   7.45   1.06   0  0.00 0.00      0.00         11.70         1.06 2007
2   408419     5123 2007-08-04 03:14 922    76.44 75.63     11.78 35.23  5.21 0.65 0.22 2.82     0 1.52   2.49   1.95   0  1.08 0.11      0.65          9.22         1.19 2007
3   407407     5123 2007-08-04 02:28 515    66.93 76.97     15.66 72.97  6.41 1.75 0.58 0.39     0 3.69   3.11   0.58   0  0.19 0.19      1.17         10.10         0.39 2007
> filter(social_data,social_data$AuthorID==116713)
  ThreadID AuthorID       Date  Time WC Analytic Clout Authentic Tone ppron    i we  you shehe they posemo negemo anx anger sad focuspast focuspresent focusfuture Year
1   404456   116713 2007-08-03 19:20 42    73.03  59.4     37.24    1  9.52  7.14  0  0.0     0 2.38   2.38   7.14   0     0   0      7.14         14.29         0 2007
2   408419   116713 2007-08-04 20:43  8    71.61  99.0      7.84   99 25.00 12.50  0 12.5     0 0.00  12.50   0.00   0     0   0     12.50          0.00         0 2007
>
```

And similarly as degree these all are simply has the highest degree and all was in the thread 408419 which is considered to be a thread that well connected the threads.

```
> diameter
[1] 3
```

```
> cat("Average Path Length: ", averagePath)
Average Path Length:  1.832402>
```

```
Order by Degree
> print(head(tabularised[order(-degree),]), digits = 3)
       degree betweenness closeness   eig
120287     15        80.5       0.05 1.000
93742      10        27.5       0.04 0.811
116713      8         8.0       0.04 0.731
5123        8        33.0       0.04 0.682
128505      6         0.0       0.03 0.657
127866      6         0.0       0.03 0.657
>
```

```
Order by Betweenness
> print(head(tabularised[order(-betweenness),]), digits = 3)
       degree betweenness closeness   eig
120287     15        80.5       0.05 1.000
5123        8        33.0       0.04 0.682
93742      10        27.5       0.04 0.811
116713      8         8.0       0.04 0.731
128505      6         0.0       0.03 0.657
53906       4         0.0       0.03 0.256
```

```
Order by Closeness
> print(head(tabularised[order(-closeness),]), digits = 3)
       degree betweenness closeness      eig
59603       1           0      1.00 0.00e+00
50158       1           0      1.00 1.52e-17
127993      1           0      1.00 0.00e+00
34292       1           0      1.00 1.52e-17
125322      3           0      0.33 6.10e-17
128400      3           0      0.33 6.10e-17
```

```
Order by Eigenvector Centrality
> print(head(tabularised[order(-eig),]), digits = 3)
       degree betweenness closeness   eig
120287     15        80.5       0.05 1.000
93742      10        27.5       0.04 0.811
116713      8         8.0       0.04 0.731
5123        8        33.0       0.04 0.682
105196      6         0.0       0.03 0.657
128505      6         0.0       0.03 0.657
```

APPENDIX

This contained the R.script that was used to analyzed and answer all of the questions.

Section Title

```
#A
#set up the working space
setwd("C:/Users/Jessica Lim/Downloads")
web <- read.csv("webforum.csv")
#Data fields
#ThreadID  Unique ID for each thread  we  "We, us, our" words
#AuthorID  Unique ID for each author  you  "You" words
#Date  Date  shehe  "She, her "him words
#Time  Time  they  "They" words
#WC  Word count of the text of the post  posemo  Expressing positive emotions
#Analytic  Summary: Analytical thinking  negemo  Expressing negative emotions
#Clout  Summary: Power, force, impact  anx  Indicating anxiety
#Authentic  Summary: Authentic tone of voice  anger  Indicating anger
#Tone  Summary: Emotional tone  sad  Indicating sadness
#ppron  "I, we, you" words  focuspast  Expressing a focus on the past
#i  "I, me, mine" words   focuspresent  Expressing a focus on the present
#focusfuture  Expressing a focus on the future  focusfuture
# Expressing a focus on the future

#A.1.
library(dplyr)
library(data.table)

#without thinking multple post made by users
web$Date <- as.Date(web$Date,"%Y-%m-%d")

month <- table(format(web$Date, "%Y-%b"))

year <- table(format(web$Date, "%Y"))
```

```
plot(year, type='l', col="blue", lwd=5, ylab = "Count", xlab = "Years", main= "Participants over years")

plot(month, type='l', col="green", lwd=5, ylab = "Count", xlab = "Months over Years", main=
"Participants over months ")


#for group by to see user trends

web_user <- web %>%

        group_by(AuthorID)


year_user <- plot(table(format(web_user$Date, "%Y")),type='b', col="blue", lwd=5, ylab = "Count", xlab
= "Years", main= "Participants over years")

year_user


web_user <- group_by(as.Date(web$Date,"%Y-%m-%d"))

month_user <- plot(table(format(web_user$Date, "%Y-%b")), type='b', col="green", lwd=5, ylab =
"Count", xlab = "Months over Years", main= "Participants over months ")

month_user



#A.2.

#https://www.statology.org/correlation-test-in-r/

#https://www.datanovia.com/en/blog/how-to-perform-multiple-t-test-in-r-for-different-variables/

#http://www.sthda.com/english/wiki/correlation-analyses-in-
r#:~:text=Correlogram%20is%20a%20graph%20of%20correlation%20matrix.%20Useful,correlation%20c
oefficients%20are%20colored%20according%20to%20the%20value.cor()


# Load required R packages

install.packages('corrplot')

library(corrplot)

library(RColorBrewer)


#remove Date and Time to see the correlation for the other linguistic variables

unique(year(web$Date))
```

```
par(oma=c(0,0,2,0), mfrow = c(1, 3))

web_cor <-cor(web %>% filter(year(Date)==2002)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2002",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2003)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2003",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2004)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2004",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2005)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2005",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2006)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2006",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2007)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2007",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2008)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2008",mar=c(0,0,2,0))
```

```
web_cor <-cor(web %>% filter(year(Date)==2009)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2009",mar=c(0,0,2,0))

web_cor <-cor(web %>% filter(year(Date)==2010)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2010",mar=c(0,0,2,0))



web_cor <-cor(web %>% filter(year(Date)==2011)%>% select(-Date,-Time))
corrplot(web_cor, type="upper", order="hclust",
        col=brewer.pal(n=8, name="RdYlBu"),title = "2011",mar=c(0,0,2,0))
#spesific p-values from t-test

library(corrplot)
web_cor_no_date_no_time <- web %>% select(-Date,-Time)

par(oma=c(0,0,2,0), mfrow = c(1, 1))

corrplot(cor(web_cor_no_date_no_time),        # Correlation matrix
        method = "shade", # Correlation plot method
        type = "full",    # Correlation plot style (also "upper" and "lower")
        diag = TRUE,      # If TRUE (default), adds the diagonal
        tl.col = "black", # Labels color
        bg = "white",     # Background color
        title = "Correlation of Linguistic Variable of Webforum in Total",       # Main title
        col = NULL, # Color palette    # Change vertical position
        mar=c(0,0,2,0))

#per variables
#Visualize the hourly pedestrian count for the four locations above, on the 31st December 2021,
```

```
#using an appropriate plot. Hint: First filter using the Date column to get the relevant data.
# Filter for date (31/12/2021) using earlier extracted data for the four
#locations
web_year =web %>% group_by(year(Date)) %>% select(-Date,-Time)

# Create line plot for comparison
#Four_Locations =DEC_long %>% filter(Sensor_Location ==
c("Melbourne.Central","Southern.Cross.Station", "Southbank",
 #                                "The.Arts.Centre"))
#Create faceted histograms
library(ggplot2)
qplot(, data = Four_Locations, geom = "histogram") +
  facet_wrap(~Sensor_Location, ncol = 4) + ggtitle("Melbourne City Hourly
Pedestrian Count Distributions")

web$Year <- format(as.Date(web$Date,format="%Y-%m-%d"),"%Y")

graph = aggregate(web[,c(5:23)], by = web[24], data = web, FUN = mean )

ggplot(data = graph, mapping = aes(x = Year, group = 1))+
  geom_line(size = 1, aes( y = WC, color = "WC"))+
  geom_line(size = 1, aes( y = Clout, color = "Clout"))+
  geom_line(size = 1, aes( y = Authentic, color = "Authentic"))+
  geom_line(size = 1, aes( y = Tone, color = "Tone"))+
  theme(axis.text.x = element_text(size = rel(1), angle = 45))+
  labs(title = "Linguistic variables VS Year Pt. 1", y = "Linguistic variables", x = "Year")


ggplot(data = graph, mapping = aes(x = Year, group = 1))+
  geom_line(size = 1, aes( y = ppron, color = "Ppron"))+
  geom_line(size = 1, aes( y = i, color = "i"))+
```

```
  geom_line(size = 1, aes( y = we, color = "we"))+
  geom_line(size = 1, aes( y = you, color = "you"))+
  geom_line(size = 1, aes( y = shehe, color = "shehe"))+
  geom_line(size = 1, aes( y = they, color = "they"))+
  geom_line(size = 1, aes( y = posemo, color = "posemo"))+
  geom_line(size = 1, aes( y = negemo, color = "negemo"))+
  geom_line(size = 1, aes( y = anx, color = "anx"))+
  geom_line(size = 1, aes( y = anger, color = "anger"))+
  geom_line(size = 1, aes( y = sad, color = "sad"))+
  geom_line(size = 1, aes( y = focuspast, color = "focuspast"))+
  geom_line(size = 1, aes( y = focuspresent, color = "focuspresent"))+
  geom_line(size = 1, aes( y = focusfuture, color = "focusfuture"))+
  theme(axis.text.x = element_text(size = rel(1), angle = 45))+
  labs(title = "Linguistic variables VS Year Pt. 2", y = "Linguistic variables", x = "Year")

web_linguistics =  aggregate(web %>% select(-ThreadID,-AuthorID,-Date,-Time,-Year),by=web$Year,
FUN=mean)

#B

web_thread <- web %>%

        group_by(ThreadID)

web_thread_posemo <- aggregate(x=web$posemo, by=list(web$ThreadID,
year(web$Date)),FUN=mean)

colnames(web_thread_posemo) <- c("ThreadID", "Year", "Mean")

web_thread_negemo <- aggregate(x=web$negemo, by=list(web$ThreadID,
year(web$Date)),FUN=mean)

colnames(web_thread_negemo) <- c("ThreadID", "Year", "Mean")

web_thread_anger <- aggregate(x=web$anger, by=list(web$ThreadID, year(web$Date)),FUN=mean)

colnames(web_thread_anger) <- c("ThreadID", "Year", "Mean")
```

```
web_thread_sad <- aggregate(x=web$sad, by=list(web$ThreadID, year(web$Date)),FUN=mean)
colnames(web_thread_sad) <- c("ThreadID", "Year", "Mean")

web_thread_anx <- aggregate(x=web$anx, by=list(web$ThreadID, year(web$Date)),FUN=mean)
colnames(web_thread_anx) <- c("ThreadID", "Year", "Mean")

install.packages('ggthemes')

library(ggthemes)

ggplot(web_thread_posemo, aes(fill=ThreadID, y=Mean, x=Year)) +
  geom_bar(position='stack', stat='identity') +
  geom_point()+
  geom_text(data = web_thread_posemo,aes(label=ThreadID),hjust=-0.2, size=3)
  theme_wsj()

posemo_data=web_thread_posemo[order(web_thread_posemo$Year,-web_thread_posemo$Mean),]
posemo_data = posemo_data[!duplicated(posemo_data$Year),]
ggplot(data = web_thread_posemo, aes(x = Year, y = Mean, color = ThreadID, group = ThreadID))+
  geom_point() +
  geom_text(data = posemo_data, aes(label = ThreadID), hjust = -0.2, size = 3) +
  labs(title = "Positive Emotion Count Over The Year",y = "Positive Emotion Mean", x = "Year")



install.packages("scatterplot3d") # Install
install.packages("rgl")
require(scatterplot3d)
require(rgl)

# Custom shapes/colors
```

```
unique(web_thread_posemo$Year)
year_colour <- c("red","blue","yellow","green","purple","brown","pink","plum2","black","khaki")
year_colour <- as.numeric(web_thread_posemo$Year)

# Plot
s3d <- scatterplot3d(web_thread_posemo[1:4722,1:3],pch=16,color = year_colour)+
  title("Happiness Indicator by posemo Mean over every ThreadID over years")

#C
#https://towardsdatascience.com/how-to-model-a-social-network-with-r-878b3a76c5a1
library(tidyverse)
library(ggraph)
library(tidygraph)



install.packages(c("igraph", "igraphdata"))
library(igraph)
library(igraphdata)

#Choose which part
social_data <- web%>% filter(between(Date, as.Date("2007-08-01"),as.Date("2007-08-05")))
social_data <- subset(social_data, social_data$AuthorID != -1)

SocialData = as.data.frame(cbind(social_data$AuthorID, social_data$ThreadID))
colnames(SocialData) = c("Author", "Thread")



UniqueAuthor = as.data.frame(unique(social_data$AuthorID))
colnames(UniqueAuthor) = c("Author")

#create graph
```

```
g <- make_empty_graph(directed = FALSE)

# add vertices
for (i in 1 : nrow(UniqueAuthor)) {
  g <- add_vertices(g, 1, name = as.character(UniqueAuthor$Author[i]))
}

# make complete graph for each club and add to g
for (k in unique(SocialData$Thread)){
  temp = SocialData[(SocialData$Thread %in% k),]
  if(nrow(temp)>1){
  # combine each pair of agents to make an edge list
  Edgelist = as.data.frame(t(combn(temp$Author,2)))
  colnames(Edgelist) = c("P1", "P2")
  # add edges
  for (i in 1 : nrow(Edgelist)) {
    g <- add_edges(g,
                   c(as.character(Edgelist$P1[i]),as.character(Edgelist$P2[i])))
  }
  }
  else {
    g = g
  }
  # following line just to check groups are correct, not needed for graph
  print(temp)
}

g = simplify(g)
plot(g)+
  title("Social Network Graph for Forum 2007-08-01 to 2007-08-05")
```

```
#C2

#https://www.r-bloggers.com/2021/04/social-network-analysis-in-r/#:~:text=Social%20Network%20Analysis%20in%20R%2C%20Social%20Network%20Analysis,and%20analyzing%20the%20structural%20properties%20of%20the%20network.

V(g)$label <- V(g)$name

V(g)$degree <- degree(g)


hist(V(g)$degree,
    col = 'green',
    main = 'Histogram of Node Degree',
    ylab = 'Frequency',
    xlab = 'Degree of Vertices')


#network diagram

set.seed(222)

plot(g,
    vertex.color = 'green',
    vertext.size = 2,
    edge.arrow.size = 0.1,
    vertex.label.cex = 0.8)


#hubs

hs <- hub_score(g)$vector

as <- authority.score(g)$vector

set.seed(123)

plot(g,
    vertex.size=hs*30,
    main = 'Hubs',
    vertex.color = rainbow(52),
    edge.arrow.size=0.1,
    layout = layout.kamada.kawai)
```

#Analysis

#https://visiblenetworklabs.com/2021/04/16/understanding-network-centrality/#:~:text=Network%20Centrality%3A%20Understanding%20Degree%2C%20Closeness%20%26%20Betweenness%20Centrality,and%20some%20are%20easier%20to%20understand%20than%20others.

#https://www.webpages.uidaho.edu/~stevel/517/RDM-slides-network-analysis-with-r.pdf

```
degree <- g %>% degree() %>% print()

degree

plot(names(degree),degree,xlab = "ThreadID", ylab = "Number of Degree") + title("Degrees of Nodes")

text(names(degree),degree,names(degree),pos=1)


closeness <- g %>% closeness() %>% round(2) %>% print()

closeness

plot(names(closeness),closeness,xlab = "ThreadID", ylab = "Closeness Level") + title("Closeness of Nodes")

text(names(closeness),closeness,names(closeness),pos=1)


betweenness <- g %>% betweenness() %>% print()

betweenness

plot(names(betweenness),betweenness,xlab = "ThreadID", ylab = "Betweeness") + title("Betweeness of Nodes")

text(names(betweenness),betweenness,names(betweenness),pos=1)


eig = as.table(evcent(g)$vector)

eig

plot(names(eig),eig,xlab = "ThreadID", ylab = "Eigenvector Centrality") + title("Eigenvector Centrality of Nodes")

text(names(eig),eig,names(eig),pos=1)


averagePath = average.path.length(g)

diameter = diameter(g)

diameter
```

```
tabularised = as.data.frame(rbind(degree, betweenness, closeness, eig))
tabularised = t(tabularised)

cat("Average Path Length: ", averagePath)

cat("\nDiameter: ", diameter, "\n\n")

# Print properties ordered by degree
cat("\nOrder by Degree\n")
print(head(tabularised[order(-degree),]), digits = 3)

# Print properties ordered by betweenness
cat("\nOrder by Betweenness\n")
print(head(tabularised[order(-betweenness),]), digits = 3)

# Print properties ordered by closeness
cat("\nOrder by Closeness\n")
print(head(tabularised[order(-closeness),]), digits = 3)

# Print properties ordered by eigenvector centrality
cat("\nOrder by Eigenvector Centrality\n")
print(head(tabularised[order(-eig),]), digits = 3)
```

JESSICA LIM 31954081

FIT 3152

ASSIGNMENT 1